

GenBank format

```
LOCUS       CMGLOAD                      1185 bp    DNA        linear   VRT 18-APR-2005
DEFINITION  Cairina moschata (duck) gene for alpha-D globin.
ACCESSION   X01831
VERSION     X01831.1  GI:62724
KEYWORDS    alpha-globin; globin.
SOURCE      Cairina moschata (Muscovy duck)
  ORGANISM  Cairina moschata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
REFERENCE   1  (bases 1 to 1185)
  AUTHORS   Erbil,C. and Niessing,J.
  TITLE     The primary structure of the duck alpha D-globin gene: an unusual
            5' splice junction sequence
  JOURNAL   EMBO J. 2 (8), 1339-1343 (1983)
  PUBMED    10872328
COMMENT     Data kindly reviewed (13-NOV-1985) by J. Niessing.
FEATURES             Location/Qualifiers
     source           1..1185
                     /organism="Cairina moschata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8855"
     CAAT_signal      20..24
     TATA_signal      69..73
     precursor_RNA    101..1114
                     /note="primary transcript"
     exon             101..234
                     /number=1
     CDS              join(143..234,387..591,939..1067)
                     /codon_start=1
                     /product="alpha D-globin"
                     /protein_id="CAA25966.2"
                     /db_xref="GI:4455876"
                     /db_xref="GOA:P02003"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="InterPro:IPR002340"
                     /db_xref="InterPro:IPR009050"
                     /db_xref="UniProt/Swiss-Prot:P02003"
                     /translation="MLTAEDKKLIVQVWEKVAGHQEEFGSEALQRMFLAYPQTKTYFP
HFDLHPGSEQVRGHGKKVAAALGNAVKSLDNLSQLSELNLHAYNLRVDPVNFKLLA
QCFQVVLAAHLGKDYSPMHAAFDKFLSAVAVLAEKYR"
     repeat_region    227..246
                     /note="direct repeat 1"
     intron           235..386
                     /number=1
     repeat_region    289..309
                     /note="direct repeat 1"
     exon             387..591
                     /number=2
     intron           592..939
                     /number=2
     exon             940..1114
                     /number=3
     polyA_signal     1095..1100
     polyA_signal     1114
ORIGIN
1  ctgcgtggcc tcagcccttc caccctcca cgctgataag ataaggccag ggcgggagcg
61  cagggtgcta taagagctcg gccccgcggg tgtctccacc acagaaaccc gtcagttgcc
121 agcctgccac gccgctgccg ccattgctgac cgccgaggac aagaagctca tcgtgcaggt
181 gtgggagaag gtggtctggc accaggagga attcggaaat gaagctctgc agaggtgtgg
241 gctgggcccc gggggcactc acaggttgag cagcaggagc caggagccct gcagcgggtg
301 tgggctggga cccagagcgc acggggtgac gggctgagat gggcaagca gcagggcacc
361 aaaactgact ggcctcgctc cggcaggatg ttctcgctct acccccagac caagacctac
421 ttcccccaat tcgacctgca tcccggctct gaacagggtc gtggccatgg caagaaagtg
481 gcggtcgccc tgggcaatgc cgtgaagagc ctggacaacc tcagccaggc cctgtctgag
541 ctcagcaacc tgcattgccta caacctgcgt gttgacctgt tcaacttcaa ggcaagcggg
601 gactagggtc cttgggtctg ggggtctgag ggtgtggggt gcagggtctg ggggtccagg
661 ggtctgagtt tcctggggtc tggcagtcct gggggctgag ggccagggtc ctgtggtctt
721 ggggtaccag gtccctgggg ccagcagcca gacagcaggg gctgggattg catctgggat
781 gtgggccaga ggotgggatt gtgtttgga tgggagctgg gcaggggcta gggccagggt
841 gggggactca gggcctcagg gggactcagg gggggactga gggagactca gggccatctg
901 tccggagcag gggactaag ccctggtttg ccttgagctg gctggcacag tgcttccagg
961 tggtgctggc cgcacacctg ggcaaagact acagccccga gatgcatgct gcctttgaca
1021 agttcttgct ccgcgtggct gccgtgctgg ctgaaaagta cagatgagcc actgcctgca
1081 cccttgacc ttcaataaag acaccattac cacagctctg tgtctgtgtg tgctgggact
1141 gggcagtcgg ggtccagggt agggctgggt tgcttcaca catcc
```

//

Header

Indeholder information ang. Organisme, publikation, Accession ID mm.

FEATURE blok

Indeholder en beskrivelse af forskellige elementer i DNA sekvensen.

CDS: **C**oding **S**equ**S**ence. Indeholder koordinater på den protein kodende del af et gen. Bemærk de tre intervaller.

ORIGIN blok

Indeholder selve DNA sekvensen.

FASTA format

Header

Hver enkelt sekvens indledes med en header linie der starter med ">". Typisk står der her et navn på sekvensen, evt. suppleret med yderligere information så som databaseID mm.

Sekvens

FASTA filer kan bruges til alle typer sekvens: DNA, RNA, Protein mm. Sekvensblokken ender ved næste ">" eller slutningen af filen, hvis det er den sidste sekvens.

>alpha-D

```
ATGCTGACCGACTCTGACAAGAAGCTGGTCCTGCAGGTGTGGGAGAAGGTGATCCGCCAC
CCAGACTGTGGAGCCGAGGCCCTGGAGAGGTGCGGGCTGAGCTTGGGGAAACCATGGGCA
AGGGGGGCGACTGGGTGGGAGCCCTACAGGGCTGCTGGGGGTTGTTTCGGCTGGGGGTGAG
CACTGACCATCCCGCTCCCGCAGCTGTTACACCTACCCCCAGACCAAGACCTACTTCC
CCCACTTCGACTTGCACCATGGCTCCGACCAGGTCCGCAACCACGGCAAGAAGGTGTTGG
CCGCCTTGGGCAACGCTGTCAAGAGCCTGGGCAACCTCAGCCAAGCCCTGTCTGACCTCA
GCGACCTGCATGCCTACAACCTGCGTGTGACCCCTGTCAACTTCAAGGCAGGCGGGGGAC
GGGGGTGAGGGGCCGGGGAGTTGGGGGCCAGGGACCTGGTTGGGGATCCGGGGCCATGCC
GGCGGTACTGAGCCCTGTTTTGCCTTGCAGCTGCTGGCGCAGTGCTTCCACGTGGTGCTG
GCCACACACCTGGGCAACGACTACACCCCGGAGGCACATGCTGCCTTCGACAAGTTCCTG
TCGGCTGTGTGCACCGTGCTGGCCGAGAAGTACAGATAA
```

>alpha-A

```
ATGGTGCTGTCTGCCAACGACAAGAGCAACGTGAAGGCCGTCTTCGGCAAAATCGGCGGC
CAGGCCGGTGACTTGGGTGGTGAAGCCCTGGAGAGGTATGTGGTCATCCGTCATTACCCC
ATCTCTTGCTGTCTGTGACTCCATCCCATCTGCCCCCATACTCTCCCCATCCATAACTG
TCCCTGTTCTATGTGGCCCTGGCTCTGTCTCATCTGTCCCCAACTGTCCCTGATTGCCTC
TGTCCCCCAGGTTGTTTCATCACCTACCCCCAGACCAAGACCTACTTCCCCCACTTCGACC
TGTACATGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTGGCGGAGGCACTGGTTG
AGGCTGCCAACCACATCGATGACATCGCTGGTGCCCTCTCCAAGCTGAGCGACCTCCACG
CCCCAAAGCTCCGTGTGGACCCCGTCAACTTCAAAGTGAGCATCTGGGAAGGGGTGACCA
GTCTGGCTCCCCTCCTGCACACACCTCTGGCTACCCCTCACCTCACCCCTTGCTCACC
ATCTCCTTTTGCCTTTCAGCTGCTGGGTCACTGCTTCCCTGGTGGTTCGTGGCCGTCCACTT
CCCCCTCTCTCCTGACCCCGGAGGTCCATGCTTCCCTGGACAAGTTCGTGTGTGCCGTGGG
CACCGTCCTTACTGCCAAGTACCGTTAA
```